

Claims

We claim:

1. For use with a search engine that processes user queries, a system that locates documents containing search words corresponding to a present user query comprising:

an index builder that stores locations of documents indexed by word in an index based on a present query-independent static rank that has been assigned to each document;

an index partitioner that orders and partitions the index into index partitions that each contain location information about a group of one or more documents having a continuous range of static ranks that is a subset of an overall range of static ranks;

an index scanner that progressively scans the index partitions starting with a partition containing those documents with the highest static rank to locate documents containing a search word; and

a scorer that calculates a score based on a present set of documents located thus far in the search and on the range of static ranks of a next partition to be scanned and wherein the index scanner scans the next partition to locate documents containing a search word if the calculated score is above a target score.

2. The system of claim 1 comprising a document ranker that assigns the static rank to a document and wherein the document ranker includes a link analyzer that detects a number of links that reference the document and wherein the static rank is a function of the number of links that reference the document.

3. The system of claim 1 wherein the document ranker includes a document usage monitor that detects a number of times the document has been accessed in response to a query and wherein the static rank is a function of the number of times the document has been accessed.

4. The system of claim 1 wherein the scorer comprises a dynamic rank computation module that computes a dynamic rank for the present set of located documents and wherein the score is based on the dynamic rank.

5. The system of claim 4 wherein the dynamic rank computation module computes the dynamic rank by totaling the number of located documents in the present set of located documents.

6. The system of claim 4 wherein the dynamic rank computation module computes the dynamic rank by determining a quality of match value for documents in the present set of located documents.

7. The system of claim 4 wherein the scorer calculates the score by adding a first weighted portion of the maximum static rank assigned to a document in the next partition to a second weighted portion of the dynamic rank.

8. The system of claim 1 wherein the index partitioner partitions the index into such that the size of a partition is smaller than or equal to a cache size.

9. The system of claim 7 wherein the first and second weighted portions are determined based on system load.

10. For use with a search engine that processes user queries, a method for locating documents containing a search word found in a present user query comprising:

- assigning a present query-independent rank to each document to be searched;
- ordering the documents to be searched in order of the assigned present query-independent rank and grouping the ordered documents to be searched into partitions by present query-independent rank;
- indexing documents in a partition by mapping a location for each document to words contained in the document to form an index;
- scanning the partitions in present query-independent rank order by iteratively i) searching a highest ranked unsearched partition for a search word found in the user query to add to a present set of located documents located thus far; ii) calculating a score based on a present set of located documents and the present query-independent rank of documents indexed in a next highest ranking unsearched partition; iii) comparing the calculated score to a target score; and iv) continuing to search the next highest ranking unsearched partition until the calculated score is higher than the target score; and
- returning search results including the document locations in the present set of located documents when the calculated score is higher than a target score.

11. The method of claim 10 comprising detecting a number of links that reference a document and assigning a static rank to the document based on the number of links.

12. The method of claim 10 comprising monitoring document usage to determine a number of times a document has been retrieved by previous queries and assigning the present query-independent rank based on the number of time a document has been retrieved.

13. The method of claim 10 comprising grouping the ordered documents into partitions having a size smaller than a size of a cache designated for storing portions of the index.

14. The method of claim 10 comprising calculating the score by computing a dynamic rank for the present set of located documents.

15. The method of claim 14 comprising computing the dynamic rank by totaling a number of located documents in the present set of located documents.

16. The method of claim 14 comprising calculating the score by adding a first weighted portion of the dynamic rank to a second weighted portion of the highest present query-independent rank for a document in the next highest ranked partition.

17. The method of claim 17 comprising determining the first weighted portion based on a search engine load level.

18. One or more computer readable media comprising computer-executable instructions for performing the method of claim 10.

19. For use with a search engine that processes user queries, One or more computer readable media comprising computer readable instructions for retrieving documents containing search words in a query by:

assigning a static rank to documents;

indexing the documents by mapping document locations to words contained in the document to construct an index;

ordering and partitioning the index by document based on the static rank assigned to the document;

iteratively searching, in static rank order, a highest ranking unsearched partition to return locations for documents containing search words in the query;

calculating a score based on a relevance of documents returned and the static rank assigned to a next partition to be searched; and continuing to search the next partition until the calculated score is higher than a target score; and

returning document locations as a query result when the calculated score exceeds the target score.

20. The one or more computer readable media of claim 19 wherein the static rank is assigned to a document based on a number of documents that reference the document.

21. The one or more computer readable media of claim 19 wherein the static rank is assigned based on a number of times the document has been returned by previous queries.

22. The one or more computer readable media of claim 19 wherein the score is calculated by calculating a dynamic rank based on the relevance of documents returned thus far in the search.

23. The one or more computer readable media of claim 22 wherein the dynamic rank is calculated based on a number of documents returned thus far in the search.

24. The one or more computer readable media of claim 22 wherein the dynamic rank is calculated by adding a first weighted portion of the dynamic rank to a second weighted portion of the static rank of a document in the next partition to be searched.

25. For use with a search engine that processes user queries, an apparatus for locating documents containing a search word found in a present user query comprising:

means for assigning a present query-independent rank to each document to be searched;

means for ordering the documents to be searched in order of the assigned present query-independent rank and grouping the ordered documents to be searched into partitions by present query-independent rank;

means for indexing documents in a partition by mapping a location for each document to words contained in the document to form an index;

means for scanning the partitions in present query-independent rank order by iteratively i) searching a highest ranked unsearched partition for a search word found in the user query to add to a present set of located documents located thus far; ii) calculating a score based on a present set of located documents and the present query-independent rank of documents indexed in a next highest ranking unsearched partition; iii) comparing the calculated score to a target score; and iv) continuing to search the next highest ranking unsearched partition until the calculated score is higher than the target score; and

means for returning search results including the document locations in the present set of located documents when the calculated score is higher than a target score.

26. The apparatus of claim 25 comprising means for detecting a number of links that reference a document and wherein the means for assigning a present query-independent rank assigns the rank to the document based on the number of links.

27. The apparatus of claim 25 comprising means for monitoring document usage to determine a number of times a document has been retrieved by previous queries and wherein the means for assigning the present query-independent rank assigns the rank based on the number of time a document has been retrieved.

28. The apparatus of claim 25 comprising means for calculating the score by computing a dynamic rank for the present set of located documents.

29. The apparatus of claim 28 wherein the means for computing the dynamic rank computes the dynamic rank by totaling a number of located documents in the present set of located documents.

30. The apparatus of claim 28 wherein the means for calculating the score calculates the score by adding a first weighted portion of the dynamic rank to a second weighted portion of the highest present query-independent rank for a document in the next highest ranked partition.